Electronic Version
Stylesheet Version v1.1.1

[System and Method for the Classification of Electronic Communication]

Abstract

From an electronic message, we extract any destinations in selectable links, and we reduce the message to a "canonical" (standard) form that we define. It minimizes the possible variability that a spammer can introduce, to produce unique copies of a message. We then make multiple hashes. These can be compared with those from messages received by different users to objectively find bulk messages. From these, we build hash tables of bulk messages and make a list of destinations from the most frequent messages. The destinations can be used in a Real time Blacklist (RBL) against links in bodies of messages. Similarly, the hash tables can be used to identify other messages as bulk or spam. Our method can be used by a message provider or group of users (where the group can do so in a p2p fashion) independently of whether any other provider or group does so. Each user can maintain a "gray list" of bulk mail senders that she subscribes to, to distinguish between wanted bulk mail and unwanted bulk mail (spam). The gray list can be used instead of a whitelist, and is far

easier for the user to maintain.